

Demultiplexing Report

Angela Crabtree

8/12/2021

Distribution of Read Pairs in Demultiplexed Bins

Inputs for the python script, `demux.py`, were 4 FASTQ files (R1, R2, R3, R4), one tab-delimited text file with barcode id and sequence information, and a user-specified q-score cutoff. The cutoff determined which reads failed to have good-quality barcode sequences and were thus relegated to the “junk” bin. Note that this cutoff does NOT apply to the sequencing coverage of the reads themselves.

Here are the command-line arguments used:

```
## ./demux_err.py
## -1 /projects/bgmp/shared/2017_sequencing/1294_S1_L008_R1_001.fastq.gz
## -2 /projects/bgmp/shared/2017_sequencing/1294_S1_L008_R2_001.fastq.gz
## -3 /projects/bgmp/shared/2017_sequencing/1294_S1_L008_R3_001.fastq.gz
## -4 /projects/bgmp/shared/2017_sequencing/1294_S1_L008_R4_001.fastq.gz
## -b /projects/bgmp/shared/2017_sequencing/indexes.txt
## -q 20
## -o REAL-output
```

If barcodes in FASTQ files did not match barcodes in the barcode info file, they were also binned in “junk”. If both forward and reverse barcodes for a read pair matched those in info file, but were not the same, read swapping occurred and read pairs were binned in “swap” bins. If barcodes matched, they were binned in the appropriate index bin. Counts were recorded and written to a csv output file.

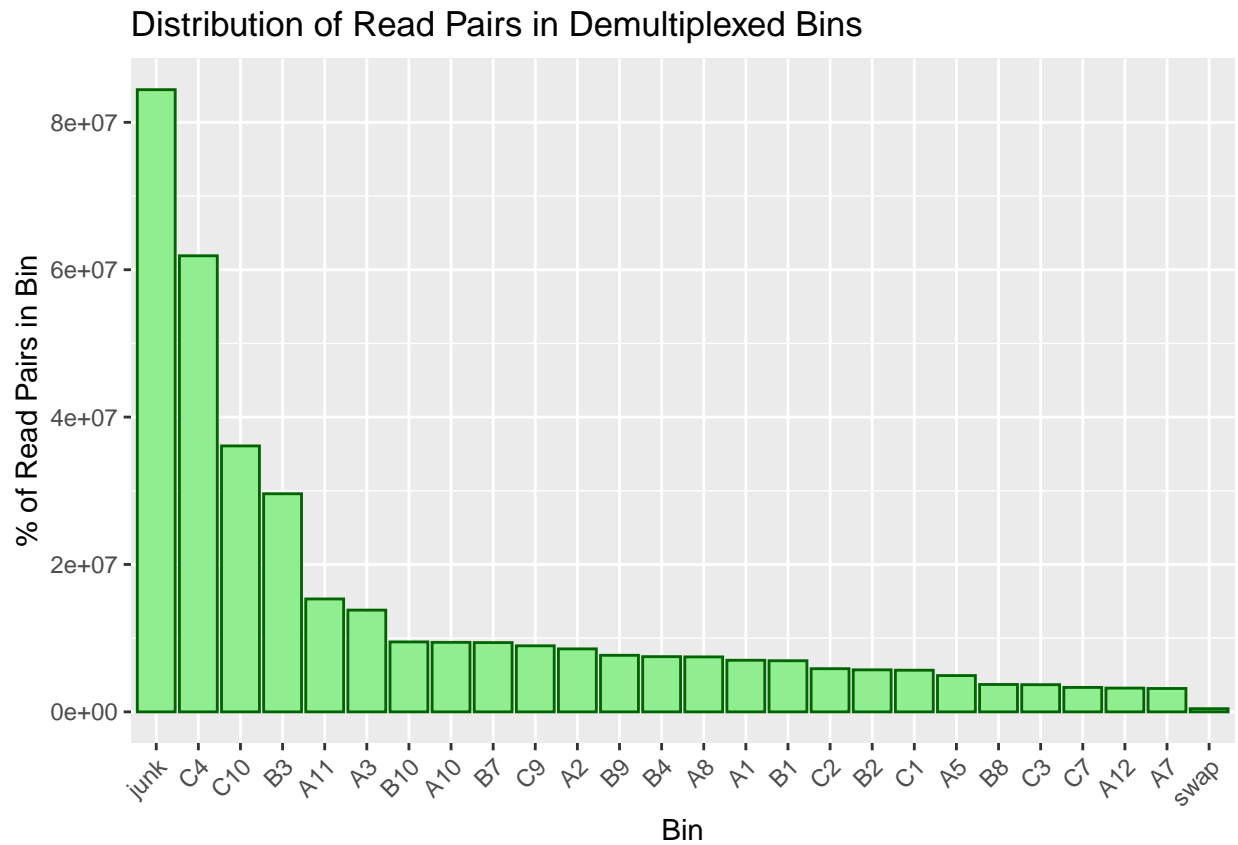
Here are the counts of read pairs binned:

##	bin	rpcount	rpp
## 1	junk	84433220	23.2440410
## 2	swap	432046	0.1189401
## 3	B1	6941570	1.9109793
## 4	A5	4926263	1.3561754
## 5	C1	5653676	1.5564286
## 6	B9	7674791	2.1128314
## 7	C9	8964532	2.4678906
## 8	C3	3688632	1.0154618
## 9	B3	29599418	8.1485710
## 10	C4	61891888	17.0385256
## 11	A11	15322832	4.2182986
## 12	C7	3318839	0.9136597
## 13	B2	5702479	1.5698638
## 14	A1	7001570	1.9274970
## 15	B7	9400229	2.5878358
## 16	A3	13809544	3.8016980

##	17	B4	7492308	2.0625947
##	18	A12	3225068	0.8878450
##	19	C10	36089262	9.9351924
##	20	A2	8546398	2.3527804
##	21	C2	5858542	1.6128272
##	22	A10	9433812	2.5970810
##	23	B8	3715385	1.0228268
##	24	A7	3174718	0.8739839
##	25	B10	9490524	2.6126935
##	26	A8	7459189	2.0534772

bin = index id or bin name; rpcount = read pair count; rpp = read pair percent of total

Distribution of Read Pairs in Demultiplexed Bins



Other Stats

Percent of Junk Reads:

23.24404 %

Percent of Swapped Reads:

0.1189401 %

Summary statistics:

##	bin	rpcount	rpp
##	A1 : 1	Min. : 432046	Min. : 0.1189
##	A10 : 1	1st Qu.: 5108116	1st Qu.: 1.4062
##	A11 : 1	Median : 7475748	Median : 2.0580
##	A12 : 1	Mean :13971028	Mean : 3.8462
##	A2 : 1	3rd Qu.: 9476346	3rd Qu.: 2.6088
##	A3 : 1	Max. :84433220	Max. :23.2440
##	(Other):20		